

RNA2DMut

RNA2DMut will generate all possible point mutations of an input sequence and predicts structural information based on the Boltzmann 2D structural ensemble. RNA2DMut_Evaluate will calculate the same data, but for a set of user defined mutants.

Input

The input for RNA2D mut consists of up to four things:

An RNA sequence (up to 600 nt long) comprised of A,G,C, or U(T). Any IUPAC base symbol is allowed, however, only AGCU(T) is evaluated in the calculation.

A constraint mask of equal length to the sequence that limits the mutations made. A "." character allows all 3 possible mutations, an "x" character forbids any mutation, while "Y" and "R" allow mutations to pYrimidines and puRines, respectively. The default value

An input structure in dot-bracket notation, where matched brackets "(" and ")" define base pairs and dots "." define single stranded nucleotides. This only affects the final output 2D structure image, but does not affect the calculation. The default output is the centroid 2D structure of the wild-type input sequence more info below.

A temperature, in degrees Celcius, for the calculation. Energy parameters can be scaled based on temperature. The default value is 37 °C, human body temperature.

Example using influenza A splicing regulatory motif

Input data:

```
Sequence:          GGUCUGAAAAAUGAUCUUCUGAAAAUUUGCAGGCCUAUCAGAAACGAAUGGGGGUGCAGAUG
Constraint Mask:   .....xYYYYYYY.....xxxx.....
2D Structure:      .....(((.....))).....((((.....((((.....)))))).....))
Temperature (degrees C): 33
```

In this example, we use a sequence for a splicing regulatory motif from influenza A. A constraint mask is applied to freeze the branch point "A" and the 4 nucleotides at the 3' splice site (AGGC) disallowing mutations (indicated with an "x" characters). The polypyrimide tract is constrained to mutate to pYrimidies (indicated with "Y" character; puRines would be constrained with an "R"). Nucleotides allowed to mutate freely are indicated with "." In the Constraint Mask.

The secondary structure to be displayed is defined in the 2D Structure entry. Here, the desired fold is indicated with a dot-bracket structure where matching "(" and ")" brackets indicated paired nucleotides, and "." dots indicate single-stranded nucleotides.

Temperature (in degrees C) is entered as 33 to simulate the cooler environment of the human lung.

Output

After an RNA2DMut run four files are generated. The first file (Fig. 1) contains data for each mutant: the mutant name (numbered sequentially from 0 [the wild-type input] to N [where N is the highest number of mutations made]), the *sequence*, the *minimum free energy (MFE) structure* in dot-bracket notation, the *Gibbs folding free energy (ΔG)* in kcal/mol, the ensemble *centroid structure* (the structure with the minimum distance to all other folds in the 2D structure ensemble), and the *ensemble diversity (ED)* metric (the average distance between structures in the 2D structure ensemble).

| 33°C Results | Sequence | MFE structure | ΔG | Centroid structure | ED |
|--------------|---|--|------------|--|-------|
| Mutant_0 | GGUCUGAAAAUGAUCUCUUGAAAAUUUGCAGGCCUUAUCAGAAACGAUUGGGGUGCAGAUG | (((.....)).....(((((((.....)))))).....)) | -17.23 | (((.....)).....(((((((.....)))))).....)) | 2.63 |
| Mutant_111 | GGUCUGAAAAUGAUCUCUUGAAAAUUUGCAGGCCUUAUCAGAAACGAUUGGGGUGCAGAUG | (((.....)).....(((((((.....)))))).....)) | -16.15 | (((.....)).....(((((((.....)))))).....)) | 19.31 |
| Mutant_99 | GGUCUGAAAAUGAUCUCUUGAAAAUUUGCAGGCCUUAUCAGAAACGAUUGGGGUGCAGAUG | (((.....)).....(((((((.....)))))).....)) | -17.23 | (((.....)).....(((((((.....)))))).....)) | 19.26 |
| Mutant_23 | GGUCUGAAAAUGAUCUCUUGAAAAUUUGCAGGCCUUAUCAGAAACGAUUGGGGUGCAGAUG | (((.....)).....(((((((.....)))))).....)) | -17.23 | (((.....)).....(((((((.....)))))).....)) | 18.94 |
| Mutant_105 | GGUCUGAAAAUGAUCUCUUGAAAAUUUGCAGGCCUUAUCAGAAACGAUUGGGGUGCAGAUG | (((.....)).....(((((((.....)))))).....)) | -17.23 | (((.....)).....(((((((.....)))))).....)) | 18.81 |
| Mutant_74 | GGUCUGAAAAUGAUCUCUUGAAAAUUUGCAGGCCUUAUCAGAAACGAUUGGGGUGCAGAUG | (((.....)).....(((((((.....)))))).....)) | -12.14 | (((.....)).....(((((((.....)))))).....)) | 18.8 |

Figure 1. Partial output file for RNA2DMut showing the WT sequence and top 5 highest ED mutants.

The second output file shows each position in the wild-type sequence, all mutations made, and their effect of the ED. An example is shown in Figure 2 below where the *nucleotide (nt) number*, *wild type (WT) base* and *WT ED* are shown, followed by the *mutant (Mut) base* and the *Mut name* (to allow referencing to the output from the first output file).

| nt | WT | WT_ED | Mut | Mut | Mut_ED | Mut | Mut | Mut_ED | Mut | Mut | Mut_ED |
|----|----|-------|-----|-----------|--------|-----|-----------|--------|-----|-----------|--------|
| 1 | G | 2.63 | A | Mutant_1 | 5.47 | C | Mutant_2 | 5.89 | U | Mutant_3 | 5.83 |
| 2 | G | 2.63 | A | Mutant_4 | 2.36 | C | Mutant_5 | 6.14 | U | Mutant_6 | 6.42 |
| 3 | U | 2.63 | A | Mutant_7 | 6.32 | C | Mutant_8 | 5.97 | G | Mutant_9 | 8.79 |
| 4 | C | 2.63 | A | Mutant_10 | 5.94 | G | Mutant_11 | 5.89 | U | Mutant_12 | 4.96 |
| 5 | U | 2.63 | A | Mutant_13 | 2.69 | C | Mutant_14 | 3.11 | G | Mutant_15 | 3.05 |
| 6 | G | 2.63 | A | Mutant_16 | 2.38 | C | Mutant_17 | 2.54 | U | Mutant_18 | 2.41 |
| 7 | A | 2.63 | G | Mutant_19 | 4.32 | C | Mutant_20 | 12.51 | U | Mutant_21 | 13.67 |
| 8 | A | 2.63 | G | Mutant_22 | 7.39 | C | Mutant_23 | 18.94 | U | Mutant_24 | 4.99 |
| 9 | A | 2.63 | G | Mutant_25 | 6.08 | C | Mutant_26 | 2.47 | U | Mutant_27 | 2.45 |
| 10 | A | 2.63 | G | Mutant_28 | 4.78 | C | Mutant_29 | 2.68 | U | Mutant_30 | 2.47 |
| 11 | A | 2.63 | G | Mutant_31 | 2.73 | C | Mutant_32 | 2.6 | U | Mutant_33 | 3.58 |
| 12 | U | 2.63 | A | Mutant_34 | 2.37 | C | Mutant_35 | 2.86 | G | Mutant_36 | 2.5 |
| 13 | G | 2.63 | A | Mutant_37 | 8.12 | C | Mutant_38 | 10.58 | U | Mutant_39 | 18.7 |
| 14 | A | 2.63 | | | | | | | | | |
| 15 | U | 2.63 | C | Mutant_40 | 2.32 | | | | | | |
| 16 | C | 2.63 | U | Mutant_41 | 6.97 | | | | | | |
| 17 | U | 2.63 | C | Mutant_42 | 3.8 | | | | | | |
| 18 | U | 2.63 | C | Mutant_43 | 2.67 | | | | | | |
| 19 | C | 2.63 | U | Mutant_44 | 2.81 | | | | | | |
| 20 | U | 2.63 | C | Mutant_45 | 2.68 | | | | | | |
| 21 | U | 2.63 | C | Mutant_46 | 2.52 | | | | | | |
| 22 | G | 2.63 | A | Mutant_47 | 2.4 | C | Mutant_48 | 2.77 | U | Mutant_49 | 2.41 |
| 23 | A | 2.63 | G | Mutant_50 | 3.45 | C | Mutant_51 | 3.45 | U | Mutant_52 | 2.42 |
| 24 | A | 2.63 | G | Mutant_53 | 4.63 | C | Mutant_54 | 7.5 | U | Mutant_55 | 4.67 |
| 25 | A | 2.63 | G | Mutant_56 | 3.22 | C | Mutant_57 | 2.84 | U | Mutant_58 | 3.16 |

Figure 2. Example second output file showing the first 25 nt of the influenza splice site motif. The branch point (nt 14) is constrained to not change, while the following 7 nts in the polypyrimidine tract are constrained to remain pyrimidines.

The next two output files are 2D structures annotated using the differences in ED between the WT and Muts. First is generated a structure showing the maximum possible increase in ED at each nt (Fig. 3).

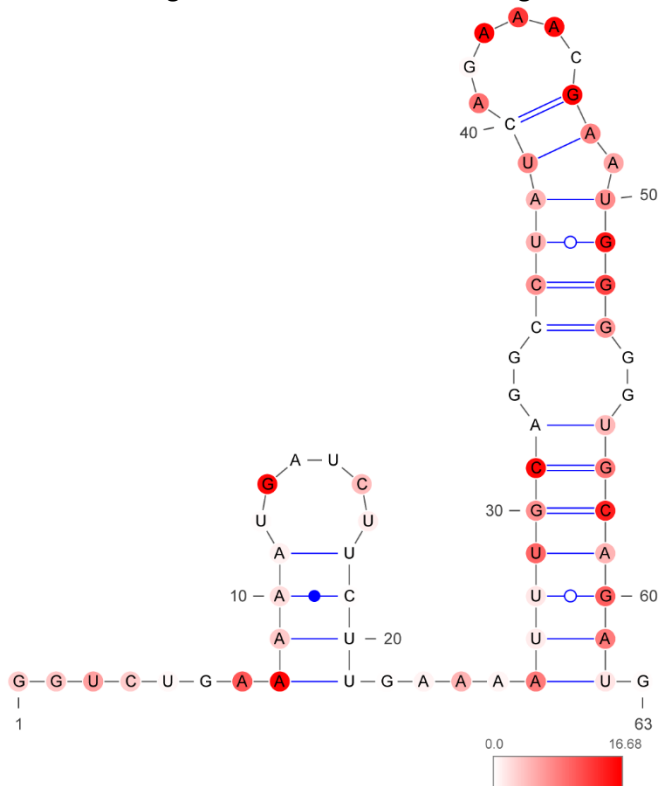


Figure 3. 2D image showing maximum possible ED change at each position. The magnitude of change in ED, (Mut - WT) is indicated by the red heat map (redder nt have higher ED in the Mut than WT).

The second image file shows the maximum possible decrease in ED going from WT to Mut (Fig. 4).

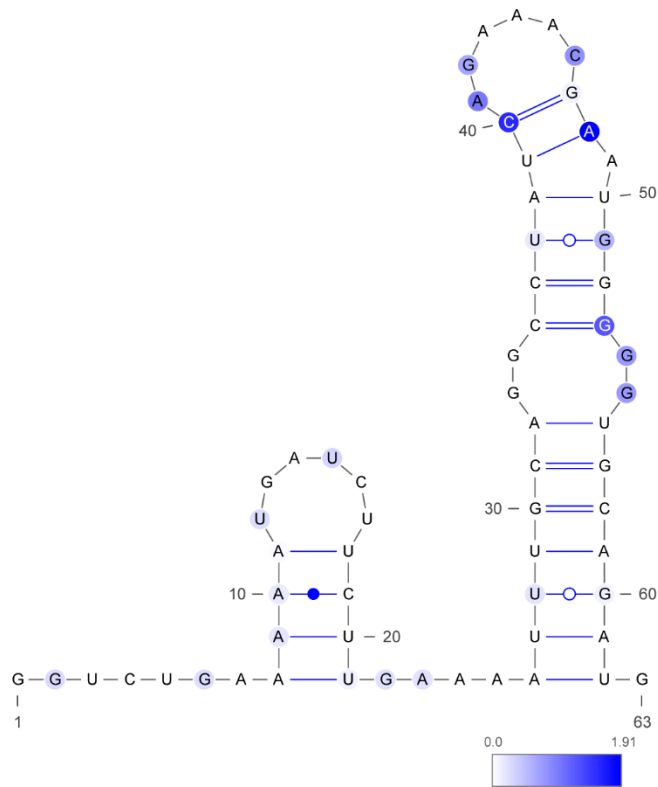


Figure 4. 2D image showing maximum possible ED change at each position. The magnitude of change in ED, (WT- Mut) is indicated by the blue intensity map (more blue nt have lower ED in the Mut than WT).

The final two output files are text files that contain the VARNA command line input needed to generate the 2D images in Figures 3 and 4.

RNA2DMut_Evaluate

The program RNA2DMut_Evaluate takes an input set of sequences in fasta format and a temperature then outputs their minimum energy folds, ensemble centroid folds and ensemble diversity. Only one output file is generated, similar in structure to the first outfile of RNA2DMut, only each line of output is for a different user defined mutation, instead of all possible point mutations.